Historical databases and the researcher

Schurer, Kevin

KEVIN SCHURER

Historical Databases and the Researcher

The title for this paper was choosen in haste. I only realised the significance of this fact later when I started to write down some ideas on the subject. I found myself asking two questions, 'what is an historical database, and who is the researcher?'. Decisions had to be made. Following some thought on the matter the conclusion was reached that an historical database should include any item of information which an historian perceives as being of use for the construction of historical fact and theory. Therefore, a database may have within it archives as housed by the P.R.O and local record offices; transcripts and research workings such as those in the library of the Society of Genealogists; manuscripts and published works of historical interest. On the other hand, the researcher can be seen to be anyone who may potentially use or add to this collection. However, these definitions are unfortunately too diverse for the context of this paper, therefore a restriction has had to be introduced. It was decided to narrow the range by including in the historical database only documentary sources available in machine-readable form[1]. The definition of researcher however remains unchanged, yet it is important to note that anyone should be able to use such a database since this qualification would exclude local data exchange networks, such as that operated by the users of Quarry Bank 1851, a computer software package for the analysis of census material[2].

The inclusion of the words 'computer' and 'historian' in the same sentence may to many appear incongrous. Historians, not entirely by accident, have often been cast as a traditional, conservative sort, not the kind of people to be dramatically influenced by recent technological inventions. Unfortunately it appears that this view is not completely ill-founded. Ten years ago a report assessing the impact of computers on the study of social science within universities concluded that history was last, a long way down the

[1] However it has ben suggested that published material in machine-readable form obtained from automated typesetting machines should also be included in such a database. See *M. Thaller's* contribution in this volume.

[2] *T. C. Lewis and G. S. Nunn:* Quarry Bank 1851: Census Data Retrieval and its application in schools, Heinemann Computers in Education, London, 1983. Longman also plan a data exchange for users of their computer package *Census Analysis.*

field, behind the other disciplines[3]. Eighty-five percent of all social science departments used a computer, yet only 26 percent of history departments did so, the nearest rival was politics with 77 percent of departments using computers. In 'real' Terms the figure of 26 percent meant that just eight history departments in England indicated that they had made use of a computer. Of individuals, only 5 percent of historians had used a machine, whereas 35 percent of all social scientists had. Education was the next lowest subject in which 24 percent of individuals had become involved with computing[4]. No doubt, in the period since this report the number of historians using computers has markedly increased. However, if a comparable report were to be conducted now, it might well reveal that the gulf between historians and their fellow social scientists had widened. Indeed, although the overall number of computer using historians has grown, with regard to the 'state-of-the-art', in some respects the situation may well have deteriorated. This point will be clarified later on. Of technological innovations affecting historical research the computer is to be found a long way down the pecking-order, behind both the development of the printing press and the invention of the photocopier[5]. As such, the greatest use to which historians put the computer is undoubtedly word-processing.

So far a dark and gloomy background to computerised historical research has been painted. To this it is hoped that a lighter image will be added, yet when viewing the picture as a whole it must be realised that it is the contrast between the two that probably forms the most important and striking aspect. One brighter feature is that there already exists a national data archive, situated at the University of Essex; set up to collect, house and distribute machine-readable datasets relevant to research in the social sciences[6]. The bulk of the historical datasets available from the archive are based on the nineteenth-century census returns, notably those from the national two percent sample of 1851, deposited by Professor Michael Anderson of Edinburgh University, but the collection also includes files of aggregated parish

---

[3] *Computing and the Social Sciences: A Report to the SSRC*, Social Science Research Council, London, 1973. See especially Appendix II, pp. 23-25.

[4] One must not, however, discount historical research that may have been carried out under the banner of another subject. Geography for example was high on the list of those possessing and teaching computer skills.

[5] See *H. J. Hanham:* 'CLIO's Weapons', in: Daedalus, Spring 1971, pp. 509-19.

[6] *SSRC Survey Archive Data Catalogue: Guide to the Survey Archive's social science data holdings and allied sciences*, University of Essex, Colchester, (no date).

register data, insurance policies and marriage registers. A brief summary of these datasets follows:

- Records of Greenwood & Batley Ltd., machine tool producers, 1836-1900
- Aggregated demographic data from parish registers, 1537-1837
- Industrial and occupational data from 1911 census, Tayside.
- Marriage register data, nine Oxford parishes, 1837-1970.
- Marriage register data for London Kentish parishes, 1851-1853, 1874-1875.
- Aggregated demographic data, nine parishes in Tendring, Essex, 1538-1838.
- Migration in Northern Highlands, 1851-1891.
- Social structure in early 20th century Belfast. (census data)
- 1851 census, Norwich Wards, 1% sample.
- 1851 census of Seaham Harbour, County Durham.
- Censuses of 1851-1871, Sheerness Naval Dockyard
- 1861 census of Southwark and Christchurch, London.
- Index to 18th century fire insurance policy registers.
- 1851 census, Co. Antrim.
- Slave Ship records, 1817-1843.
- Slave Ship trade, 1791-1799 (from House of Lords survey).
- 1851 census, 2% national sample.
- Suffolk census data, 1851-1871.
- Ship tax for Harwich, Essex.
- Scottish Poor Law Commission statistics, 1844.
- British investors in the 16th and 17th centuries.
- Sample of insurance policies, 1750-1850.

Anybody can deposit data with the archive, indeed one family history society has already done so[7]. Probably more important is the fact that the ESRC requires that any data generated as a result of their funding should be deposited with the archive. However, unfortunately the situation is not as good as this requirement suggests. Those depositing data with the Essex Data Survey Archive, funded by the ESRC or not, form a clear minority of computer-using historians. The computerised datasets from many research projects, particularly prost-graduate thesis work, have no doubt been deleted or lie disused in various departments throughout the country. It seems that

[7] The West Surrey Family History Society have deposited 1861 census data for Southwark, (Christchurch and St. Saviours).

most people, once having spent so much time and effort on collecting and making machine-readable a series of historical documents, wish to either safeguard the information from other people or gain something in return for their labours. Even for manual work, few of the many census indexes made by family historians have been deposited at either the Society of Genealogists or local record offices[8]. Ironically although the photocopying of documentary sources can raise considerable problems due to copyright, historians can make machine-readable or hand transcripts of them without any liability. Perhaps archivists should insist that copies of any transcriptions must be placed with the appropriate record office and/or other relevant data depository[9].

It is easy to simply say 'send all machine-readable transcripts and data files to a data archive', yet the practicabilities of the situation are not quite so straightforward. The Essex Data Archive is already under-staffed and under-financed and may well not entirely welcome a flood of computerised transcripts. Even if the archive did have the facilities to handle such an influx the question, 'would people actually use these data files?' must be asked. There certainly seems to be a reluctance for academics to re-work 'second-hand' data and the degree to which local and family historians would want to use computerised data rather than selected print-outs of the information they require is questionable. What good is a library if no-one wants to read or borrow the books? This issue is of particular importance if the chief criterion for the funding of such an archive is the quantity of users rather than the quantity of depositors and quality of data. The question of data-quality leads on to what is probably the most significant problem area: how usable are the data files? The question of utility is very much interwoven with the whole issue of standards which has dominated this discussion[10]. However, it is not intended to rehearse the issues surrounding the standards debate in this article.

Historians in further education, as has been noted already, are no great users of computers, they probably understand the amorphous boxes even less. Worse still, the advice that many historians are offered shows little respect for the integrity of the historical source material. It is often assumed that a software package such as the much used Statistical Package for the Social Sciences (SPSS) will meet all of the historians computer requirements.

---

[8] *J. Gibson and C. Chapman (eds.):* Census Indexes and Indexing, Federation of Family History Societies, 1981.

[9] *B. Collins:* The computer as a research tool, in: Journal of the Society of Archivists, 7 (1), 1982, pp. 6-12.

[10] See especially Computers in Genealogy, vol. 1, no. 1 and vol. 1, no. 7.

In terms of the data file manipulation and statistical calculations available from SPSS this assumption may well be true, particularly in the case of the recently improved version[11]. However, the problem is that the traditional way of preparing information for this type of packaged analysis is to divide the data up into eighty character length records consisting of a number of numerically coded variables or fields[12]. Thus the inherent logic of the data is sacrificed to an obsolete computer technology. It is in this sense that the situation of historical computer-based research is possibly worse now than it was ten years ago. For example, of the few computer users a decade ago, several had acquired various programming and data-manipulation skills[13], whereas today many historian computer users approach their research with the view that if it cannot be done within SPSS then it cannot be done at all. It is clear that such an approach is unsatisfactory.

Attempts have been, and are being made to remedy the situation and a framework for data analysis is being constructed. A clear distinction has to be made between data input and data collection. Data should be collected replicating the layout of the source as closely as possible, with logical records or units of variable length and secondary fields or data elements also of variable length[14]. The stage of data processing in which truncation, and coding occurs should be separated from data collection, with all processing being undertaken inside rather than outside the computer[15]. For documents of a standard format such as census returns, tithe awards and post-1812 parish registers, it is relatively easy to design a format which will represent the logical units of the original document in machine-readable form. A format has already been suggested elsewehere for the first of these documents[16].

---

[11] *N. H. Nie et al.:* SPSS: Statistical Package for the Social Sciences, New York, 1975. *Anon.:* SPSS-X User's Guide, Chicago, Ill.: SPSS Inc., 1983.

[12] The main text book on the subject, although now very much out of date, recommends the use of numeric coding and restricting the record length to 80 columns. See *E. Shorter:* The Historian and the Computer, New Jersey, 1971.

[13] *R. Schofield:* English Historians and the Computer, in: Historical Methods Newsletter, 7, 1974, pp. 111-114.

[14] *R. Schofield and R. Davies:* Towards a flexible Data Input and Record Managment System, in: Historical Methods Newsletter, 7, 1974, pp. 115-124.

[15] R. Floud: An Introduction to Quantitative Methods for Historians, (2nd ed.), London, 1979. See especially pp. 202-210.

[16] *K. Schurer:* Methodology: Recording Data from the Original Sources, in: Historical Social Sciences Newsletter, 2, 1984, pp. 8 -11 and *K. Schurer:* Census Enumerators' Returns and the Computer, in: Local Historian, forthcoming.

Machine-readable burial register (Bradfield, Essex 1814)

```
50/BRADFIELD/ESSEX/ONE THOUSAND EIGHT HUNDRED AND FOURTEEN
60/17/JOHN,SHIPLEY/BRADFIELD/APRIL 17TH/28/HY THOMPSON VICAR
60/18/-,BLYTH/MANNINGTREE/APRIL 24TH/4 YE/HY THOMPSON
60/19/JOHN,SEABORN/BRADFIELD/APRIL 30TH/63/HY THOMPSON
60/20/S-,FOX/BRADFIELD/JULY 1ST/76/HY THOMPSON
60/21/MARY,TURNER/BRADFIELD/JULY 6TH/69/HY THOMPSON
60/22/MARY,MORGAN/BRADFIELD/AUGUST 24TH/INFT/HY THOMPSON
60/23/S-,GOYMER/BRADFIELD/SEPTR 11TH/39/HY THOMPSON
```

Example 1

while the other sources may be collected according to the formats illustrated in the following examples. A printed burial register may be recorded in a simple form shown in example 1. Lines tagged 50 indicate the start of a new page and lines tagged 60 relate to the individual row entries. The columns across this row are indicated by slashes (/), and logical subdivisions within these units such as between prenames and surnames are shown by a comma.

A similar framework can be applied to a tithe award, as in example 2.

The lines tagged 59 and 69 indicate the column headings applicable to the fields in the lines tagged 50 and 60 respectively. Each parcel of land starts with a line tagged 50 and ends with a subsequent line 50 or a line tagged 70, which totals the quantity and amounts payable for the particular holding. Separate land divisions within a single holding (i.e. fields) are detailed on lines tagged 60. Columns with no information are represented by a single dash (-) unless they are trailing, in which case they are omitted.

However, the interests of historians are not confined to documents of a standard form and logical consistency; wills, deeds and even early parish registers all have little or no coherent logical structure. Faced with this predicament computer using historians have often adapted the data to fit their programmes by juxtaposition, abbreviation and retention of only the information that interests them; thus rendering the dataset near worthless for the purposes of archival use. A much wiser approach to this problem would be to collect the data in a free-format fashion replicating the source, change the programmes and let the computer carry out any adaptation, retaining at all time a machine-readable copy of the source which can always be referred to or re-worked as necessary. Unfortunately, computers cannot readily understand text typed into them verbatim, in order to produce the sort of analyses historians require the computer has to be given a sense of location. This

Machine-readable tithe award (Steeple, Essex 1839)*

```
59/LANDOWNERS/OCCUPIERS
.
.
69/NO/NAME/STATE/QUANTITY/PAYABLE TO VICAR/PAYABLE TO J K +
HUNT & ELIZ HUNT & THOMAS HUNT
.
.
50/EXORS OF THE LATE ISAAC,BROWN/JOHN,RADLEY
60/12/KINGS 6A/ARABLE/6,2,22/-/1,15,10
60/13/ROADFIELD/DO/6,2,9/-/1,17,-
60/14/6 1"/"2A/DO/6,3,14/-/1,18,6
60/15/4 A/DO/3,3,34/-/1,3,1
60/16/3 A/DO/3,3,37/-/1,3,5
60/17/4 A/DO/4,-,28/-/1,1,6
60/18/BARNFIELD/DO/3,3,5/-/1,1,6
60/19/GARDENFIELD/DO/3,1,16/-/,18,3
60/20/HOMESTEAD/DO/,2,3
60/21/2 A/GRASS/2,3,25/-/-,3,-
60/22/6 A/ARABLE/6,1,4/-/1,13,2
60/23/5A MARSH/DO/5,1,37/-/1,6,9
60/24/3A MARSH/GRASS/4,-,38/-/-,5,-
70/58,2,32/-/14,7,-
50/WILLIAM,BLAKE/GEORGE,PATTISON
60/202/LOST FIELD/ARABLE/13,2,15/1,6,-/3,-,3
50/,SOCIETY OF QUAKERS/,SOCIETY OF QUAKERS
60/218/BURIAL GROUND/GRASS/,,31
```

* Note that slash in 6 1/2 A (No 14) has been surrounded by quotation marks to distinguish it from a slash separating data elements. Also lines longer than 80 characters are continued on a subsequent line and the continuation indicated by a plus sign (+) on the continued line and by indenting the continuation line two spaces. Dashes (-) in the amounts payable and sizes of fields are permissible since these cannot be confused with dashes indicating missing data elements as these are always between two slashes.

Example 2

may be achieved by inserting a series of unique flags or pointers into the text indicating the presence of words or phrases which may be considered by the researcher to form logical data elements or variables[17]. For example, an early parish register may be represented in machine-readable for as in example 3; baptisms are recorded on lines tagged 61, burials on lines tagged 62, marriages on lines tagged 63 and prenames, surnames and dates are all flagged, flags starting and ending with an asterisk.

Alternatively, according to the structure of a particular document, the researcher may wish to interleave lines of free-field and fixed-field formats. For example, in the baptism register of St. James, Clerkenwell the incumbent generally made the entries in the format as follows: month; date; prename; " 's' or 'd' of"; father's prename; parents surname; "&"; mother's prename; "his wife; born"; date. Therefore, lines entered in this standard fashion may be tagged to indicate that they are in this fixed format, whilst lines breaking this convention can be tagged and flagged accordingly as in example 4.

However, although in this example slashes separating logical elements are substituted by spaces, such practise is potentially dangerous since if the recording of a middle name or some other detail went unnoticed, the extra space would throw the computer into disarray.

Unfortunately flagging of data elements is not a straightforward task. If all that is required for retrieval purposes is the flagging of names and occupations then the necessary flags can be quite simple. However, if the researcher wishes to link data elements and retain the context in which the data elements occur, then the process of flagging can become extremely complicated. This latter point is illustrated by work carried out by Alan Macfarlane and colleagues in a project aimed at making machine-readable every document over the period 1550 to 1750 relating to the Essex village of Earls Colne [18]. Wishing to retain the grammatical syntax of the text, the documents were broken down into a series of entities based round a subject matter. These were then flagged, bracketed and linked to other entities by nesting and numbering of brackets [19]. As example 5 shows, for a parish register the scheme

---

[17] See *M. Overton:* Computer analysis of an inconsistent data source: the case of probate inventories, in: Journal of Historical Geography, 3 (4), 1977, pp. 317 - 326 and *G. A. Dobbert:* An On-line System for Processing Loosely Structured Records, in: Historical Methods, 15 (1), 1982, pp. 16-22

[18] The research project is based at the Department of Social Anthropology, University of Cambridge.

[19] *C. J. Jardine and A. D. J. Macfarlane:* Computer Input of Historical Records for Multi-Source Record Linkage, in: *M. W. Flinn (ed.):* Proceedings of the 7th

Machine-readable Parish Register (Bradfield, Essex 1738)*

```
50/1738
61/BAP: *P WILLIAM*, THE SON OF *P SAMUEL* AND *P MARTHA +
*S CARRINGTON*, WAS BAPTIZED THE *D 4TH OF SEPTEMBER* .  +
1738
61/BAP: *P SARAH* AND *P MARY*, DAUGHTERS OF *P THOMAS* AND +
*P BRIDGET* *S BRASSTREE* WERE BAPTIZED THE *D 8TH OF +
OCTOBER* .
62/BUR: *P SARAH* & *P MARY *S BRASSTREE* WERE BURIED THE +
*D 16TH OF OCTOBER* THE AFFIDAVIT FOR BURYING IN WOOLLEN +
REGISTERED
61/BAP: *P ELIZABETH* , THE DAUGHTER OF *P JOHN* AND *P +
SUSANNA*S ROWLAND* WAS BAPTIZED THE *D 22D OF OCTOBER* .  +
1738
61/BAP: *P JOSHUA* , THE SON OF *P JOHN* AND *P ANNE*S NUN* +
, WAS BAPTIZED THE *D 5TH OF NOVEMBER* .  1738
62/BUR: *P SARAH*S YELL* WAS BURIED THE *D 14TH OF NOVEMBER* +
.A.R.
62/BUR: *P SUSANNA*S PLUMMER* WAS BURIED THE *D 15TH OF +
NOVEMBER* .A.R.
63/MAR: *P THOMAS*S GOSS* AND *P ELIZABETH*S KING* WERE +
MARRIED THE *D 16TH
OF NOVEMBER* , THEIR BANNS BEING FIRST THRICE DULY PUBLISHED
62/BUR: *P ALICE*S KING* WAS BURIED THE *D 17TH OF DECEMBER* +
.  A.REGIST. BEING BROUGHT WITHIN EIGHT DAYS, THE TIME +
PRESCRIBED BY THE ACT OF PARLIAMENT, MADE FOR BURYING IN +
WOOLLEN.
```

* Note that in this example an asterisk is used to both end and start a flag. Lines longer than 80 characters are continued on a subsequent line and the continuation indicated by a plus sign (+) on the continued line and by indenting the continuation line two spaces.

*P = Prename

*S = Surname

*D = Date

Example 3

International Economic History Congress, Edinburgh 1978, Edinburgh, 2, 1978, pp. 71-78.

Mixed free and fixed formats (St James, Clerkenwell 1698)*

```
60/JUNE 16 MARY D. OF WILL. COOKE & MARY HIS WIFE; BORN 24 MAY
60/JUNE 19 MARY D. OF JAMES NEWTON & ELIANOR HIS WIFE; BORN 6
60/JUNE 24 ELIZ. D. OF JOHN STORER & ELIZ. HIS WIFE; BORN 24
60/JUNE 24 RICH. S. OR ROBT SAVAGE & FRAN. HIS WIFE; BORN 20
60/JUNE 26 WILL. S. OF JOHN MUNDAY & ANN HIS WIFE; BORN 13
60/JUNE 26 WILL. S. OF JOHN JENKINSON & ELIZA. HIS WIFE; +
BORN 13
60/JUNE 26 SARAH D. OF RICH. SNOW & ANN HIS WIFE; BORN 18
70/*D JUNE 27* *CN CATH.* D. OF *FN WILL.* & *MN CATHR*
*S SMITH* , BUT I BELEEVE THIS IS ILEGITIMATE
60/JUNE 28 WILL. S. OF THO. ORETON & ANN HIS WIFE; BORN 10
60/JUNE 30 WILL. S. OF WILL. PAGE & SARAH HIS WIFE; BORN 29
60/JULY 3 JOSEPH S. OF JOHN BARRETT & SARAH HIS WIFE; BORN +
20 JUNE
60/JULY 3 THOMAS S. OF ISAAC SUFFOLKE & ELIZ. HIS WIFE; BORN +
29 JUNE
60/JULY 3 WILL. S. OF WILL. DRAKE & CATHR HIS WIFE; BORN +
23 JUNE
60/JULY 5 SARAH D. OF DANIEL KENEDAY & JONE HIS WIFE; BORN 5
70/*D JULY 15* *CN THO.* S. OF MR *FN DEUEL* *S PEAD* , +
GENT., & *MN SARAH* HIS WIFE; *BD BORN 4*
60/JULY 17 SAMLL S. OF JOSHUA ATKINSON & FRAN. HIS WIFE;
BORN 14
60/JULY 17 DOROTHEA D. OF HENRY DOWNER & ELIZ. HIS WIFE;
BORN 5
60/JULY 21 JOHN S. OF ROBT NOBLE & SARAH HIS WIFE; BORN 21
60/JULY 22 ELIZ. D. OF JOHN KINGSTON & ELIZ. HIS WIFE; BORN 13
```

* Example taken from *R. Hovenden (ed):* A True Register of all the Christeninges, Mariages and Burialles in the Parishe of St James, Clarkenwell, From the Yeare of our Lorde God 1551, Vol 1 Christenings, 1551 to 1700, The Harleian Society, (Registers, (Volume IX)), London, 1884, p.376 .

*D = Date (of event)

*CN = Christian Name

*FN = Father's Name

*MN = Mother's Name

*S = Surname

*BD = Birth Date

Example 4

Example of linked flagging (from Earls Colne, Essex)*

A: Parish Register

Source text:

```
JOHN THE SON OF HENRY ABBOTT WAS BAPTISED 5TH MAY 1607
```

Logical structure:

[A person] [who has a name] [and who is involved in a kinship relation] [with another person] [who has a name]. [The first person is involved in an event] [on a date].

Input data:

```
(P (N JOHN )(K THE SON OF (P (N HENRY ABBOTT ) ) )
(E WAS BAPTISED (D 5TH MAT 1607 ) ) )
```

B: Will

Source text:

```
HENRY ABBOTT AND JONE HIS WIFE DO CLAIM FOR HOLD A TENEMENT
IN CHURCH STREET
```

Input data:

```
(P *1 (N HENRY ABBOTT ) ) AND (P (N JONE ) (K (1 HIS ) WIFE )
(H DO CLAIM FOR HOLD (L A TENEMENT IN CHURCH STREET ) )
```

* P - Person
N - Name
K - Kinship relation
E - Event
D - Date
H - Landholding
L - Description of land

Example taken from *C.J.Jardine and A.D.J.Macfarlane:* Computer input of Historical Records for Multi-Source Record Linkage, In: *M.W.Flinn (ed):* Proceedings of the 7th International Economic History Congress, Edinburgh 1978, 2, Edinburgh, 1978, pp.71-78

Example 5

is relatively easy to implement, however, a lengthy will may require dozens of nested and linked brackets which can only be added after the document has been read, understood, broken down into its composite entitities, and the entities linked with each other.

Regardless of the simplicity or complexity of the flags and pointers used by a researcher it is important that a number of points are adherred to. The combination of characters used as flags obviously have to be unique and must not occur in the text of the document. Yet equally a flag should not be too long, its meaning should be implicit, perhaps mnemonic, and it must be portable. Therefore, it must not contain characters that cannot be reproduced by other computers, printers or typewriters. Also it must be remembered that flags are only used as pointers to words or phrases that the current researcher feels may be of interest. No set of predetermined pointers will be absolute, other researchers will wish to sub-divide singularly flagged data elements and re-define or re-classify various flagged data elements altogether. Therefore, if data are to be stored in a data archive it is of crucial importance that any flags used do not change the shape of the data and that the original text of the document is always retrievable. It is for these reasons that in the case of free-format documents it is probably best to initially type in the text with appropriate tags indicating the structure of the document, and then add any required flags at a secondary editing stage, thus retaining the flag-free version for reference and archive purposes. Such a policy would also satisfy the demands of the many non-computer orientated historians who just want a legible, understandable line-printer copy of the data. Additionally, if those persisting with manual research prepared their transcriptions in line with the conventions used in the first of these stages, then high quality transcriptions could be fed into a computer via an optical character reader, such as a Kurzweil Data Entry Machine (KDEM) and then flagged and formatted as required.

Many of the points that have been made look slightly to the future, to a situation where data can be referred to, deposited with, and taken out of an archive in much the same way as we currently use libraries. Nothing has been said about the host of technical problems surrounding the accomplishment of this situation, about the problems of incompatibility of storage media and storage formats. In the past changes in technology have adjusted the nature and intensity of many of these problems and will, no doubt, continue to do so in the future. For example, an interesting recent innovation has been the development of the multi-format floppy disc copying micro-computer, called the 'magic machine' for short, which, capable of handling over seventy different storage formats, has been installed at the Essex Data Archive to offer a disc transfer service to those who are unable to transfer information between micro-computers at their own institutions. However, regardless of the technical abilities, such a service would be dramatically devalued if the

data being transferred could not be historically interpreted as a result of data reduction or subscription in the form of various codes and classifications.

Still looking towards the future words of caution have to be sounded not only about historical sources being converted into an historical database, but also about the data that is being created now for the historian of the future. To what use will the historian put the machine-readable documents that our society is currently creating? Will he be able to use them at all? Already the discs on which the 1960 American census data are stored cannot be read because the disc drives are no longer made and all of the old ones have been scrapped. Similarly the magnetic tapes holding the British 1961 census have already decayed to the point that researchers wanting special tabulations from them have been refused for "technical reasons". What of more everyday archives; the accounts of small-businesses, the wage-books of companies and the records of schools and local government? Will the historian of the future be able to read such files on his computer, or will he have to use a specialized machine, or will all the files he wishes to consult have been written-over as they become out-of-date anyway? What will the scientific or literary historian make of films containing no drafts, just mathematical formulae and prose, neatly edited with all trace of initial, superceded ideas deleted? All of these questions demand an answer which needs to be considered very carefully.